XML RELATED DATA EXCHANGE FROM THE TEST MACHINE TO A WEB-ENABLED MAT-DB

M Nagy, HH Over*, E Wolfart

Joint Research Centre of the European Commission – Institute for Energy, Petten, The Netherlands
Email: miklos.nagy@jrc.nl; hans-helmut.over@jrc.nl; erik.wolfart@jrc.it

ABSTRACT

*One of the JRC Petten tasks is to support European Research and Development (R&D) projects in the material and energy related areas with the management and dissemination of research results. By using XML technology test data can be entered directly from the test machines into the Web-enabled 'Materials Database' (**Mat-DB**), which is an integral part of the On-line Data Information Network (ODIN On-line Data Information Network, 2004) that has been developed at JRC Petten. Test data, which are kept in XML format and sent by R&D project partners via the World Wide Web to the Petten Server are stored within the **Mat-DB** XML module. There they can be checked and updated on-line before they are uploaded into the database. After validation by the source administrator they can immediately be retrieved and evaluated by all project partners. A pilot test with the new XML related data exchange module from test machine into **Mat-DB** has currently been started within the European 'TMF Standard' R&D project.*

Keywords: On-line materials Database, Test data, steels & alloys, Thermo-physical & mechanical properties, XML data exchange, Knowledge sharing and dissemination, Ontology

1. INTRODUCTION

The engineering sciences place considerable reliance on electronic systems to produce, store and process experimentally measured material data. The use of XML standards is already a very promising step into the next generation of the Web, often referred to as the Semantic Web. It is aimed at machine-processable information that will enable true interoperability between sort specific databases, which are different in structure and language. Data on the Semantic Web is characterized by semantic metadata called ontologies that facilitates sharing and interoperability between systems and languages. It is our aim for the future to realize this step and thus to enable JRC Petten to build up a European network for collecting and disseminating publicly available experimentally measured materials data within national and European R&D projects using the Web-enabled **Mat-DB**. Further more down the line it would allow the exchange of materials data on an international level. One of the JRC Petten data management and dissemination tasks is the establishment of a European network for collecting and disseminating publicly-available experimentally-measured materials data within national and European R&D projects in the materials and energy related areas. JRC has been developing two material databases for safeguarding and managing its experimental materials data resulting from in-house research some 20 years ago (Hurst, Kröckel, Over & Vannson, 1988; Kröckel & Over, 1994) called Alloys-DB and Corrosion-DB. Alloys-DB covers data on the mechanical and thermo-physical properties of engineering alloys at low, elevated and high temperatures for base materials and joints (Over, Rantala, Taylor & Dietz, 2003). It includes irradiation materials testing in the field of fusion and fission, tests on thermal barrier coating for gas turbines and mechanical properties testing on a corroded specimen. The Corrosion-DB refers to weight gain/loss data of high temperature-exposed engineering alloys, ceramics and hot isostatic pressed powder materials and covers corrosion tests such as oxidation, sulfidation and nitridation. The extension to other types of corrosion is under consideration. Alloys-DB and Corrosion-DB have been merged into Mat-DB. The most promising and efficient data storage method is the implementation of a direct link between the test machine, the post-processing software of a test machine and a database respectively.

Test data, which are logged in XML format and sent by project partners via the World Wide Web to the Petten Server can be stored within the Mat-DB XML database module. There it can be checked and updated on-line before it is uploaded into the database. After validation by the source administrator data can immediately be retrieved and evaluated by all project partners. A pilot test with the new XML-related data-exchange module from the test machine into **Mat-DB** has just been started within the European 'TMF Standard' (Hähner & Bressers, 2002) R&D project. This project creates huge amounts of data; each of the 20 partners performs 20 thermo-mechanical fatigue tests. Each thermo-mechanical fatigue test contains information about sources, materials, specimens and test conditions, which has already resulted in major sets of metadata. However the test results themselves contain test specific information, data for up to 100 cycles and up to 300 curve data for each cycle.

2. XML AND SEMANTIC INTEROPERABILITY

XML - eXtended Markup Language is accepted as the emerging standard for data exchange on the Web, aimed at solving the problem of interpretation and syntactic interoperability. However still not meaningful enough to describe information sources at the semantic level. One reason why semantic interoperability became an important issue is that there are currently over 300 XML e-Business formats, essentially amounting to domain specific taxonomies. As envisioned the Semantic Web is the next generation of the World Wide Web which will provide machine processable information to intelligent agents or software applications in order to help humans find their way through the huge amount of data available on this media. Knowledge representation researchers have already developed different ontology languages to represent machine-readable and understandable web languages based on current exchange standards such as XML. One of the latest proposals of the W3C is OWL (OWL Web Ontology Language, 2004), which became a W3C recommendation in 2003 as a web ontology language. This approach assumes that the information available on the web is expressed or annotated with these languages, which means that concepts are expressed by means of classes, properties and the relationships between them. Scientific databases containing vast numbers of experimental data became Web enabled in order to facilitate better knowledge sharing and reuse between the scientific communities. Although these databases are accessible, the seamless data exchange between differently standardised databases is still an unsolved problem in spite of the fact that different XML based languages were defined by the different scientific communities e.g. Markup language for the Exchange of Materials Property Data (MatML) in the field of materials science to facilitate a standardized XML-based data exchange. This solution solved a number of interoperability issues but makes the assumption that both parties have agreed on the syntax of the data exchange. This assumption fails when one searched for existing experimental data on the web since neither the syntax nor the semantics of the requested data is known before the submission of the query. The problem is that different research institutions and companies use different standards and naming conventions in their logical data model for the same data, additionally these data models are not always even accessible on the web. Hence a vast number of experimental data are remaining inaccessible or unanalyzed.

As different communities defined their own eXtended Markup Language (XML) schema by using e.g. MatML for the Exchange of Materials Property Data, femML for Finite Element Modeling Markup Language, CML for Chemical Markup Language and MathML for Mathematical Markup Language to resolve interpretation and interoperability issues during data exchange via the World Wide Web a number of problems from the data exchange perspective have been resolved. From the computation perspective a tag like <Material> carries the same semantic information as tag <H1>. The computer just simply does not know what the material tag means and how the material tag as a concept relates to another concepts like data source. From this perspective information between tags can help humans to predict what information can be found between the tags but for an XML processor this information is completely meaningless. To address the semantic interoperability problem Tim Berners Lee (Berners-Lee, Hendler & Lassila, 2001; Berners-Lee & Fischetti, 1999) the founder of the World Wide Web proposed different levels of information that can be expressed by the language on the web (Figure 1).

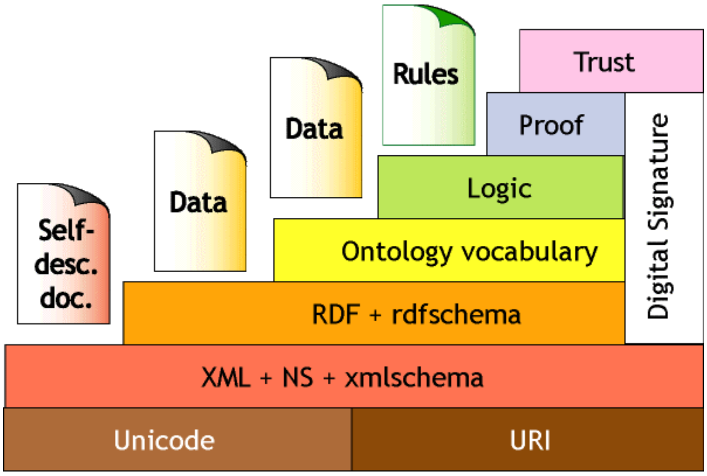


Figure 1. Different levels of information on the web expressed by the language

The functions of the different layers are as follows:

- XML, Namespaces and Schema provide a basic format for structured documents, with no particular semantics.
- Resource Description Framework (RDF) and RDF Schema is a basic assertion model that allows an entity-relationship-like model to be created for the data and allows the document structure to be constrained for predictable computable processing.
- The Ontology layer provides more powerful schema concepts, such as inverse, transitivity, uniqueness of properties that allows a system to recognise different identifiers and their relations that are talking about the same thing.
- The Logical layer turns a limited declarative language into a language with inference and functions.
- The Proof layer enables different parties to exchange assertions between each other, together with the inference path to that assertion.
- Web of trust: The proof layer together with the digital signature turns the web into the Web of trust.

To date the industrial and commercial implementations of data exchange on the World Wide Web can be placed into the second layer where the XML Schema defines XML data. Research projects from different fields have resulted in a number of prototype systems that make use of the higher layers namely ontologies to resolve semantic heterogeneity between different data sources.

For true semantic interoperability, which means sharing knowledge and information between different applications a shared set of terms describing the application domain with a common understanding is necessary. More flexibility could be achieved if not just a set of terms but the relations between them is defined. In this way applications are at least partially aware of the domain, which increases flexibility considerably.

To address the above-mentioned challenges, building on the XML and XML Schema more expressive ontology languages have been created that represent the conceptualization of the domain that explicitly representing the meaning of terms in vocabularies and relationships between those terms. They also describe a vocabulary for describing properties and classes: between others, relationships between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, the richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes. These extensions of the markup language make it possible not only to read the data but to interpret and infer about the data. This will enable the development of applications that can autonomously retrieve and manipulate the information available on the Semantic Web. While the complete realisation of Tim Berner's Lee view seems far away, achievements that have been reached so far are worth utilizing to enrich data-exchange with semantic information.

3. ARCHITECTURAL ASPECTS

Web services based on XML messaging technologies, such as Simple Object Access Protocol (SOAP) and Web Service Description Language (WSDL) are the backbone for our pilot application, which includes SOA (Service-Oriented Architecture) and Oracle XML database that provides a high-performance, native XML storage and retrieval technology. Our current implementation runs under Oracle 9i and Oracle 9iAS. Due to performance issues related to the Oracle XDB, the JRC Petten Data Management and Dissemination (DMD) sector is migrating to Oracle 10g, which can provide improved performance in XML storage and manipulation. From the knowledge engineering perspective, the Protégé ontology & knowledge base editor, which provides a user friendly and easy way to create OWL files, has been used to constructs ontologies. Since the Semantic Web is still in its infancy, which implies that a well-developed ontology engine is still non existent, the DMD is working on a ontology engine that is based on pre-defined domain specific rules that can support the XML based data-entry process.

4. MAT-DB CONTENT

The pilot study has been carried out for data exchange between test machines and the Mat-DB that stores experimentally measured mechanical and physical properties data of engineering alloys produced by European R&D projects. It covers the materials behavior at low, elevated and high temperatures for base materials and joints and also includes irradiation materials testing in the field of fusion and fission, and tests on thermal barrier coating for gas turbines. Its emphasis is on data from standardised tests and on evaluation methods, which are well established and widely accepted. Mat-DB has more than 130 entities with more than 1850 attributes. A detailed description of Web enabled retrieval and evaluation can be found in Dietz, Over & Wolfart, (2004) and Over, Wolfart & Veragten, (2004).

5. MAT-DB-ONTOLOGY

Ontology as defined in the knowledge-sharing context is the specification of the concepts of a specific domain. The integration of information resources in the different science disciplines is one of the most challenging problems that informatics faces nowadays. Domain experts and engineers are flooded with a wide variety of information sources in a wide variety of formats, ranging from raw lab instrument data, textual documents, structured and semi-structured data or even the combination of the afore-mentioned formats. Ontologies can serve as a main solution when one needs to integrate, share and exchange heterogeneous data or information in a semantically consistent way. Using ontologies for data integration has several advantages:

- A rich predefined vocabulary: the conceptual interface is independent of the different schemas or metadata.
- A comprehensive data dictionary: sufficient support for translating between different representations.
- Sufficient support for detecting inconsistent data even in the case of large data sets.

The basic structure of the Mat-DB ontology is depicted in Figure 2. Development of the Mat-DB ontology is in progress and will be extended for all the entities that Mat-DB cover. The ontology corresponds to the Mat-DB XML Schema that extends the MatML Schema with concepts like specimen and test condition, which is not described by MatML. The main purpose of creating such a metadata repository is to represent the semantic concept mappings between the different research partners and/or external clients without imposing a standardized syntax structure on them. The main advantage of this solution (Lehti & Frankhauser, 2004) is to let different institutions use their own well-established naming structures but facilitate the sharing of data by translating between the different naming conventions without explicitly hard coding it into software applications.

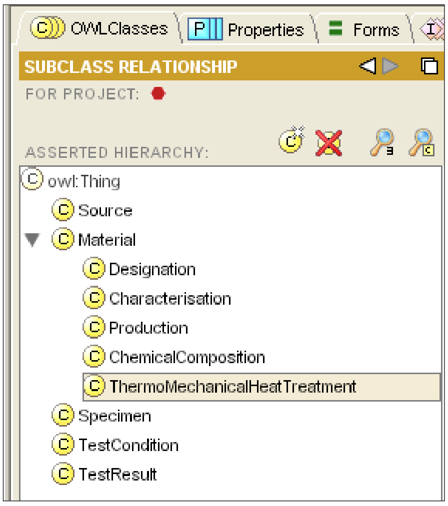


Figure 2. Mat-DB ontology class hierarchy

6. XML DATA IN MAT-DB

The high-level system architecture of the pilot XML upload study that has been carried out for uploading complex thermo-mechanical fatigue tests into the Mat-DB is shown on Figure 3. The scope of the test types will be extended for all tests those Mat-DB covers. Research partners post process the raw test results supplied by the test machine and then create XML files from it that conforms to their naming structure, which has been registered in the OWL ontology repository. Based on the mappings in the ontology repository the data file can be transferred into the Alloy DB XML module through Web-services where the data are temporarily stored before they are uploaded into the relational database. Keeping the XML files in a temporary area enables research partners to check, update and validate the data set on-line and share the data in XML format between each other where, based on the ontology repository one can resolve the structural and naming differences between formats in cases where a standardised format has not been agreed on.

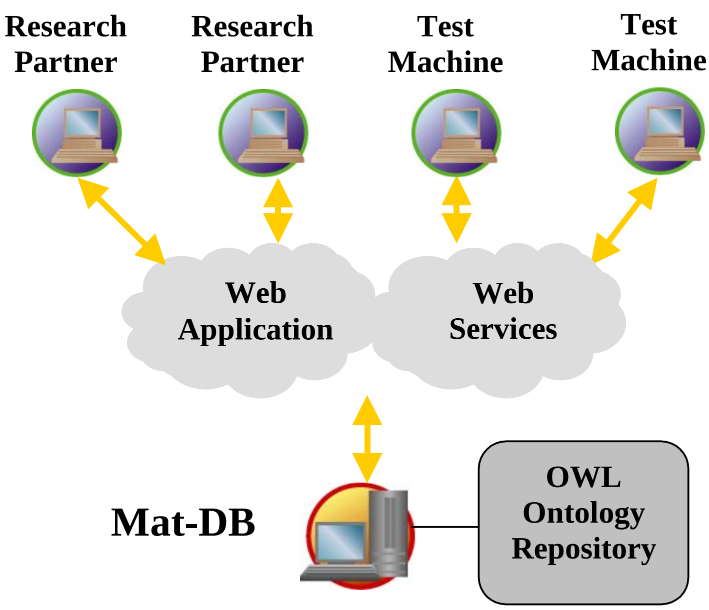
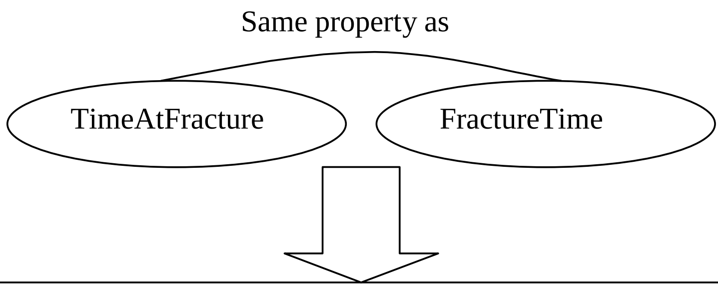


Figure 3. System architecture of the Mat-DB XML module

Since OWL is based on description logic it is possible to use a reasoner to determine inconsistencies in the ontology. However this has proved difficult in practice because of the lack of a mature ontology to reasoner language converters. In the Mat-DB ontology, besides the concept and property descriptions, different instances represent the different research partners. In our pilot study we are currently able to resolve the following differences:

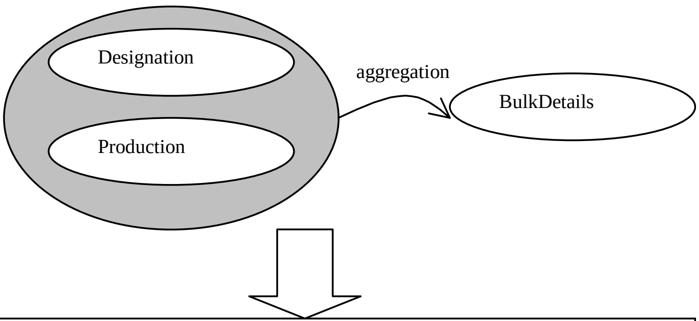
- Concept and attribute naming
- XML Structure and linguistic

Figure 4 shows that different vocabularies that describes similar attributes can be mapped directly into the OWL <owl:equivalentProperty> construct while different concepts can be mapped into <owl:equivalentClass>.




```
<owl:DatatypeProperty rdf:ID="TimeAtFracture">
   <owl:equivalentProperty rdf:resource="#FractureTime"/>
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

Figure 4. Concept and attribute similarity in OWL




```
<owl:Class rdf:ID="BulkDetails">
  <rdfs:label xml:lang="en">BulkDetails</rdfs:label>
  <rdfs:label xml:lang="de">HauptAngaben</rdfs:label>
    <owl:unionOf rdf:parseType="Collection">
      <owl:Class rdf:about="#Production"/>
      <owl:Class rdf:about="#Designation"/>
    </owl:unionOf>
</owl:Class>
```

Figure 5. Structural similarities in OWL

As depicted in Figure 5 the structural differences can be mapped into complex OWL class constructs that support basic set operations such as <owl:unionOf>. The <rdfs:label> entry provides a solution for the linguistic differences but does not affect the logical interpretation of the ontology.

7. UPLOADING DATA FROM THE TEST MACHINE TO THE MAT-DB

Scientific data servers, and in particular the ones publicly available on the Web, usually provide information retrieval techniques to access data. Uploading experimental data is a very important factor in data management. Raw data produced by the test machine can be easily converted into the XML format by a post processor and uploaded directly to the database. However in practice this process is not as straightforward as described above. Metadata needs to be created, which can then be used to mediate between different the XML formats and standards used by the different institutions.
An overview of the uploading process (Figure 6) from the test machine to the Mat-DB is as follows and is depicted in Figure 6.

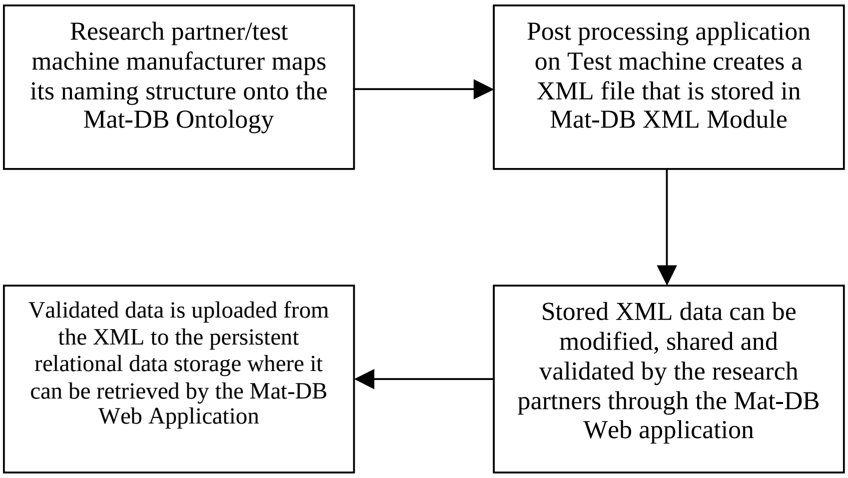


Figure 6. Flowchart of the XML Data upload

Research partners and the Data Management and Dissemination group create mappings between the different naming conventions, which can cover structurally and semantically different expressions and include them into the Mat-DB OWL ontology, e.g.:

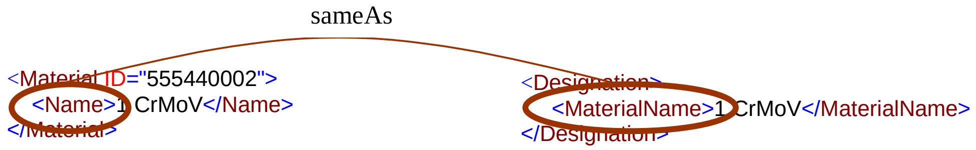


Figure 7. Attribute similarity

Instead of carrying out hard coded decision-making the post processing application can make use of the Mat-DB OWL ontology to resolve semantic differences between the raw data files or source XML files and the Mat-DB XML Schema. As a result of these process the XML file can be stored in the Mat-DB XML module. Mat-DB Web application provides a user-friendly interface to add, modify, validate and share the data set stored in XML format. In order to avoid the possibility of typical human errors like mistyping important field values (e.g. test standard, fracture mode) the Web application maintains a thesaurus list for the specific entities in the Mat-DB that helps to guarantee that the field values will not be entered incorrectly. Once the mandatory fields have been filled and an authorized person has validated the data set, the content of the XML file can be uploaded into the relational database where it is immediately available for Web enabled retrieval and evaluation (Dietz, Over & Wolfart, 2004; Over, Wolfart & Veragten, 2004). In addition to the existing database analysis routines, JRC is currently integrating **Fitit** (Fitit, 2004), - proprietary software of the Fraunhofer Institute in Germany designed for the calculation of complex material models. **Fitit** (Fitit, 2004) can be used for models defined as a set of differential equations and analytic functions, e.g. Chaboche and Kachanov. Data selected within the web-enabled Mat-DB can be sent to **Fitit** (Fitit, 2004), which then fits the data to a selected model. Once the model parameters have been calculated, they are returned to Mat-DB and the user can apply them to Finite Element (FE) lifetime calculations of high temperature exposed components with commercial codes like ABAQUS or ANSYS (Mohrmann, Denner & Hollstein, 2003) , which improves safety and reliability and saves costs.

8. CONCLUSION

In this paper we investigated the use of ontologies for materials data exchange between research partners in the European Research Projects. The results are relatively modest compared with the full potential of using ontologies to improve data discovery and exchange in a distributed and heterogeneous environment. Using a standardised ontology between material databases and from machine to database for facilitating data exchange is just the first step toward achieving semantic interoperability between heterogeneous data sources and fully in line with our future vision of using semantically rich data sets in the scientific domain. We argue that using only standardized schema-based metadata repositories will solve today's interoperability problems in the scientific community. Moving towards using distributed ontologies and mappings for each data provider is a challenging task and an active research area that will contribute to our future research interests. According to our view in future the main challenge in data interoperability and exchange is to address the problem of semantic heterogeneity in the Materials community in the creation of an industry wide ontology repository. Besides these challenges it has to be acknowledged that moving forward in this direction requires caution since the majority of usable practical semantic web-related technologies are in their infancies and the area itself is an active research topic among computer science and knowledge representation researchers. The biggest shortcoming that we identified is the OWL reasoner connection and the lack of a reliable ontology engine that can make use of the information coded in the metadata.

9. REFERENCES:

Berners-Lee T., Hendler J. & Lassila O. (2001) The Semantic Web. *Scientific American*, *184*, 34-43.

Berners-Lee T & Fischetti, M. (1999) *Weaving the web: The original design and ultimate destiny of the World Wide Web by it's inventor*. San Francisco, US: Harper

Dietz, W., Over, H. H. & Wolfart, E. (2004) Web-enabled Materials Databases of the European Commission, a Tool to Manage Experimentally Measured Data. *Materialsweek 2004*. Munich, Germany.

Fitit (2004) Homepage of Fitit. Available from: https://www.fitit.fraunhofer.de/

Hähner, P. & Bressers, J. (2002) Thermo-mechanical fatigue: the route to standardization. *Materials at High Temperature*,*19*(4),211-218.

Hurst, R.C. ,Kröckel, H. ,Over, H.H. & Vannson, P. (1988) The Use of Models in Materials Databanks. *Proc. Materials 88 - Materials and Engineering Design*. London, United Kingdom.

Kröckel, H. & Over, H. H. (1994) The Need for an European Approach to Materials Characterisation and Data Management. *Proc. International Symposium on Advanced Materials for Lightweight Structures*. Noordwijk, Netherlands.

Lehti, P. & Frankhauser, P (2004) XML Data Integration with OWL: Experiences & Challenges. *Proc. Symposium on Applications and the Internet (SAINT 2004)*, Tokyo, Japan.

Mohrmann, R., Denner, V. & Hollstein, T. (2003) Zur Lebensdauervorhersage von Austenit-Ferrit-Mischverbindungen. *Proc. Tagungsband FDBR Teil 4*, Düsseldorf, Germany.

ODIN On-line Data Information Network (2004) Homepage of ODIN. Available from: https://odin.jrc.nl

Over, H.H., Rantala, J. H., Taylor, N. & Dietz, W. (2003) A materials properties database approach to cover welded joints. *Proc. 2nd International Conference on Integrity of High Temperature Welds*. London, United Kingdom.

Over, H.H., Wolfart, E. & Veragten, E. (2004) Data Retrieval and Evaluation within the Web-enabled Alloys-DB. *Proc. 19th International CODATA Conference*. Berlin, Germany.

OWL Web Ontology Language (2004) Homepage of OWL. Available from: http://www.w3.org/TR/owl-ref/